Advance Publication

# INDUSTRIAL HEALTH

Received: September 14, 2021

Accepted: November 14, 2021

J-STAGE Advance Published Date: November 22, 2021

**Article type:** Original Article

**Title:** Inter-observer agreement and accuracy in classifying radiographs for pneumoconiosis among Asian physicians taking AIR Pneumo certification examination

**Authors:**

Naw Awn J-P[1], Agus Dwi SUSANTO[2], Erlang SAMOEDRO[2], Muchtaruddin MANSYUR[3], Sutarat TUNGSAGUNWATTANA[4], Saijai LERTROJANAPUNYA[4], Ponglada SUBHANNACHART[4], Somkiat SIRIRUTTANAPRUK[5], Narongpon DUMAVIBHAT[6], Eduardo ALGRANTI[7], John E. PARKER[8], Kurt G. HERING[9], Hitomi KANAYAMA[10], Taro TAMURA[11], Yukinori KUSAKA[12, 13] and Narufumi SUGANUMA[1]

**Affiliation:**

[1]Department of Environmental Medicine, Kochi Medical School, Kochi University, Japan

[2]Department of Pulmonology and Respiratory Medicine, Faculty of Medicine, Universitas Indonesia, Persahabatan Hospital, Indonesia

[3]Department of Community Medicine, Faculty of Medicine, Universitas Indonesia & Southeast Asian Ministers of Education Regional Centre for Food and Nutrition (SEAMEO RECFON), Indonesia

[4]Department of radiology, Central Chest Institute of Thailand, Department of Medical Services, Ministry of Public Health, Thailand

[5]Department of Disease Control, Ministry of Public Health, Thailand

[6]Department of Preventive and Social Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand

[7]Division of Applied Research, FUNDACENTRO, Brasil

[8]Pulmonary and Critical Care Medicine, Robert C. Byrd Health Sciences Center, School of Medicine, West Virginia University, USA

[9]Department of Diagnostic Radiology, Radio-oncology and Nuclear Medicine, Radiological Clinic, Miner's Hospital. Klinikum-Westfalen (Knappschaftskrankenhaus), Germany

[10]Division of Environmental Health, Department of International Social and Health Sciences, Faculty of Medical Sciences, University of Fukui, Japan

[11]Fukui City Public Health Center, Japan

[12]School of Medical Sciences, University of Fukui, Japan

[13]Kochi Medical School, Kochi University, Japan

**Corresponding Author:** Narufumi Suganuma

Department of Environmental Medicine, Kochi Medical School, Kochi University

Kohasu, Oko, Nankoku, Kochi, 783-8505 Japan.

Phone & Fax: +81-88-880-2407

E-mail: nsuganuma@kochi-u.ac.jp

ORCID: 0000-0003-1610-6216

**Running title:** CLASSIFYING PNEUMOCONIOSIS BY AIR PNEUMO READERS

**Abstract**: This study examined inter-observer agreement and diagnostic accuracy in classifying radiographs for pneumoconiosis among Asian physicians taking the AIR Pneumo examination. We compared agreement and diagnostic accuracy for parenchymal and pleural lesions across residing countries, specialty training, and work experience using data on 93 physicians. Physicians demonstrated fair to good agreement with kappa values 0.30 (95% CI: 0.20–0.40), 0.29 (95% CI: 0.23–0.36), 0.59 (95% CI: 0.52–0.67), and 0.65 (95% CI: 0.55–0.74) in classifying pleural plaques, small opacity shapes, small opacity profusion, and large opacities, respectively. Kappa values among Asian countries ranging from 0.25 to 0.55 (pleural plaques), 0.47 to 0.73 (small opacity profusion), and 0.55 to 0.69 (large opacity size). The median Youden's J index (interquartile range) for classifying pleural plaque, small opacity, and large opacity was 61.1 (25.5), 76.8 (29.3), and 88.9 (23.3), respectively. Radiologists and recent graduates showed superior performance than other groups regarding agreement and accuracy in classifying all types of lesions. In conclusion, Asian physicians taking the AIR Pneumo examination were better at classifying parenchymal lesions than pleural plaques using the ILO classification. The degree of agreement and accuracy was different among countries and was associated with background specialty training.

**Keywords:** AIR Pneumo, Chest radiograph, Diagnostic accuracy, Occupational health, Pneumoconiosis, Reader agreement

## INTRODUCTION

Pneumoconiosis, a diffuse lung disease caused by inhaled industrial or environmental dust, presents radiographically with multiple reticular or variable-sized nodular opacities[1]. Pleural plaque, an irregular, circumscribed area of dense, firm, fibrous tissue, usually resulting from asbestos exposure, appears radiographically as discrete areas of pleural thickening[2]. Screening for lung or pleural changes in a dust-exposed worker is performed primarily by periodic reviews of chest radiographs[3]. The detection and interpretation of the two conditions in a chest radiograph is highly subjective and reader-dependent. To standardize reports and facilitate international comparison of data, the International Labour Office developed a classification system (ILO classification)[4]. This classification system is composed of guidelines and a set of standard radiographs, exemplifying the spectrum of the disease. The ILO published the first edition in 1950 and made several revisions to clarify ambiguities in earlier editions but preserved the basic structure of the system. Since its establishment, the ILO classification is increasingly being adopted internationally for use in epidemiological research, screening, and surveillance of pneumoconiosis.

Screening and surveillance programs are very effective at detecting new cases of pneumoconiosis and also provide information about trend and burden of disease in workers exposed to mineral dust[5]. To promote the efficiency of screening programs in developing countries, the Asian Intensive Reader of Pneumoconiosis (AIR Pneumo) provides training and examination programs for raising physicians who can perform the ILO classification[6]. At the end of 2019, more than five hundred physicians had received training since the program began in 2006. The participating physicians have expertise in general medicine, occupational medicine, public health, pulmonology, and radiology. They include physicians from several developing Asian countries who were practicing in hospitals or working in corporations, government institutions and ministries. Most importantly, they have been working on pneumoconiosis screening.

Despite using the ILO classification, substantial variation in the interpretation of radiographs for pneumoconiosis exists among physicians[7, 8)]. Thus, before sharing epidemiological information, it is worth understanding the extent of inter-observer agreement and diagnostic accuracy among physicians of Asian countries. Therefore, the objective of this study was to examine the degree of observer agreement, diagnostic accuracy, and possible causes for reader variability in classifying radiographs for pneumoconiosis using reading results of Asian physicians taking the AIR Pneumo examinations.

## SUBJECTS AND METHODS

### *AIR Pneumo's examination film set*

The AIR Pneumo's examination film set is composed of 60 chest radiographs; the diagnosis of each radiograph was established by a panel of experts formed by 12 National Institute for Occupational Safety and Health of the United States (NIOSH) certified B Readers. The technical quality of the radiographs was classified by the 12 B Readers as ILO grade 1 (Good) or 2 (Acceptable, with no technical defect likely to impair classification of the radiograph for pneumoconiosis)[4)]. The 60-film set includes 20 radiographs with no reticular or nodular lesions, 9 boundary cases (ILO profusion classification 0/1 or 1/0), and 31 radiographs with small opacities (ILO profusion classification 1/1 or higher). Among the radiographs with small opacities, 20 have purely rounded while 4 have purely irregular opacities. Of the 31 radiographs with small opacities, 9 also have varying sizes of large opacities (opacities with the longest diameter larger than 1 cm). Nine of the 60 examination films have pleural plaques with or without calcification. Details of the AIR Pneumo's training program, development of training materials (including chest radiographs), examination, and scoring system have been published previously[9, 10)].

***Physicians' information and radiograph reading data***

Our study used 5,580 readings of 93 physicians from the two examinations conducted in Thailand (December 2018) and Indonesia (February 2019). They had taken the examination after completing an intensive 2-day AIR Pneumo training course. Physicians' information, including residing country, specialty training, and work experience, was collected through self-administered questionnaires. During the examinations, physicians independently read the chest radiographs on a standard view box in a comfortable reading room (controllable lighting with no direct sunlight) and reported the findings on reading sheets according to the ILO classification. They were given three hours to classify 60 radiographs. Each radiograph was graded for technical quality. Small opacities were classified according to their shape (rounded or irregular), size (size up to 1.5mm, 1.5–3 mm, or 3–10 mm), location (upper, middle, or lower lung zones), and profusion. Profusion was determined by side-by-side comparison with ILO standard radiographs and classified on a twelve-point scale with increasing order of concentration (codify as 0/− to 3/+ within four major profusion classifications: 0, 1, 2, and 3). Large opacities were classified as size A, B, or C, corresponding to size up to 5 cm, up to right upper lung zone, or exceeding right upper lung zone. The presence or absence of pleural plaques, their extent and width if any were recorded. We extracted data on the profusion and shape classifications of small opacities. We also obtained the size classifications of large opacities and the presence or absence of pleural plaques. Classifications on the size and location of small opacities and the technical quality of radiographs were not the purpose of this study.

***Statistical analysis***

We grouped physicians according to their residing country, specialty training, and experiences. Considering the number of years required to develop medical experience or to enroll in specialty training, years after graduation was grouped as “5 or fewer years”, “6 to 10 years”, or “11 or more years”. Information on the total number of reviewed pneumoconiosis chest radiographs, the participating physicians have encountered since they became physicians, was collected as “none”, “less than 10”, “10 to 50”, or “50 or more”. For small opacity profusion, we examined inter-observer agreement on four major profusion classifications as they showed a close correlation to the clinical severity of “normal,” “mild,” “moderate”, or “severe” conditions[11)]. When computing agreement on small opacity shape, we used only the data of 40 radiographs, i.e., 9 boundary cases and 31 radiographs with small opacities. For the other analyses, we used data of all 60 radiographs. We used a Stata module ‘kappaetc’ to compute inter-observer agreement in physicians overall and each group formed by residing country, specialty training, or experience[12)]. This command can handle any number of observers and any number of categories. It calculates the agreement coefficient by averaging the observed agreement over all pairs of observers. It also provides seven prerecorded weights, suitable for any level of measurement. We computed weighted Fleiss’ kappa to quantify the degree of agreement in classifying small opacity profusion and large opacity size and unweighted Fleiss’ kappa for agreement on small opacity shape and the presence or absence of pleural plaques[13)]. The result was interpreted values <0.2 as poor agreement, 0.21–0.4 as fair, 0.41–0.6 as moderate, 0.61–0.8 as good, and 0.81–1.0 as almost perfect agreement. Accuracy, in this study, was the ability to discriminate between normal and abnormal radiographs, i.e., the ability to classify a radiograph for the presence or absence of small opacities, large opacities, or pleural plaque; the true condition for each chest radiograph was determined based on the reading results of expert panel. Accuracy of the physicians was assessed by using only the chest radiographs that were in complete agreement for the presence or absence of small opacities, large opacities, or pleural plaque by all expert B Readers. There were 31 radiographs with and 20 radiographs without small opacities; 9 radiographs with and 41 radiographs without large opacities; 9 radiographs with and 30 radiographs without pleural plaques. A classification of 1/0 or higher profusion and any of the size classifications for large opacity by the physicians was considered as identification of small opacities and large opacities, respectively. We examined the accuracy of each physician group by plotting receiver operating characteristic (ROC) curves and computing area under the curves (AUC) against experts' diagnosis as a reference standard. An ROC curve that plots sensitivity against 1-specificity allows visual inspection of the discriminating power, while AUC quantifies the power with a value of 1.0 representing perfect discriminatory ability and 0.5 being at chance level[14)]. We used Stata's 'roccomp' command to execute ROC analysis. Assuming sensitivity and specificity are equally important in identifying each type of lesion, we calculated Youden's J index (i.e., sensitivity + specificity – 1) as a global measure of accuracy for every physician[15)]; multiplying the index by one hundred generated accuracy scores. For the accuracy score for small opacity shape classification, we computed percent agreement with the reading results of expert panel. There were 20 radiographs with purely rounded and 4 with purely irregular opacities. We then compared the accuracy scores between physician groups using the Kruskal-Wallis test with Bonferroni correction for multiple comparisons. All analyses were performed using Stata/MP 15.1 software (StataCorp., College Station, TX, USA). This study was approved by the institutional review board of Kochi Medical School (approval number: 31-68). Written informed consent from the participating physicians was waived, but opt-out consent was obtained via e-mails instead.

**RESULTS**

Table 1 presents information about our physicians. Information on specialty training and experiences (years after graduation and the number of reviewed pneumoconiosis chest radiographs) were not reported by some participating physicians. Physicians resided in India, Indonesia, Malaysia, and Thailand. They had expertise in occupational medicine, public health, respiratory health, and radiology. Specialties' representation was uneven between countries. Working duration since medical graduation ranged from 1 to 34 years. Eighteen percent of our physicians reported they had never seen a pneumoconiosis chest radiograph, while 44% encountered less than ten in their work.

(insert Table 1)

Table 2 presents the kappa values for classifying chest radiographs by physicians overall and by the groups studied. Physicians showed fair to good agreement with kappa values 0.30 (95% CI: 0.20–0.40), 0.29 (95% CI: 0.23–0.36), 0.59 (95% CI: 0.52–0.67), and 0.65 (95% CI: 0.55–0.74), respectively for classifying pleural plaques, small opacity shapes, small opacity profusion, and large opacities. The degree of agreement was different among physician groups. Physicians from Country 4, or groups formed by physicians who received radiology training, or were five or fewer years working after graduation, achieved the highest agreement in all types of lesion.

(insert Table 2)

Fig. 1 depicts the ROC curves and average AUC values of the physician groups for each pneumoconiotic lesion. Table 3 compares physician groups for their accuracy scores. Accuracy in identifying small opacities, large opacities, and the pleural plaques, as determined by AUC and accuracy scores, was different among physician groups. Physicians from Country 4, or with radiology training, or who were five or fewer years working after graduation, showed the highest accuracy (Fig. 1 and Table 3). Accuracy scores for small opacity shape classification showed a similar pattern of differences (Table 3). No substantial difference in accuracy was detected between groups formed by the reported number of reviewed pneumoconiosis chest radiographs (Table 3).

(insert Fig. 1)

(insert Table 3)

## DISCUSSION

To our knowledge, this study is the first in comparing inter-observer agreement and accuracy in classifying radiographs for pneumoconiotic lesions using the ILO classification among physicians from different Asian countries. We observed that the degree of inter-observer agreement and diagnostic accuracy varied with the observer's characteristics, namely, residing country, specialty training, and time after graduation.

Physicians in this study showed better agreement in classifying parenchymal lesions than pleural plaques using the ILO classification. However, they agreed on the shape of small opacities poorly. The degree of agreement varied between countries, with kappa values ranging from 0.47 to 0.73 (moderate to good agreement) on the distribution of small opacity profusion, from 0.55 to 0.69 (moderate to good agreement) for large opacity size, from 0.25 to 0.55 (fair to moderate agreement) for the presence or absence of pleural plaques, and 0.18 to 0.56 (poor to moderate agreement) for small opacity shape classification. The poor agreement between observers for the shape of small opacities was not unexpected. We have noted that of the 40 radiographs with small opacities from the AIR Pneumo examination film set, the expert panel agreed on small opacity shapes in only 24 radiographs. Moreover, studies that examined the shape classification of small opacities reported substantial variation existing between observers[16, 17]. Not many studies have examined inter-observer agreement involving multiple readers using the ILO classification. One Japanese study[18], which examined inter-observer agreement between film-screen radiography and two digital systems, reported the kappa values for the distribution of small opacity profusion on a twelve-point scale ranging from 0.55 to 0.64. However, their study involved a relatively small number of subject radiographs (n=30) and readers (n=3). In an American trial where seven B Readers classified 172 coal workers' chest radiographs, the reported kappa value of 0.58 for agreement on small opacity profusion was within the range of our results[19)]. In a German study, seven physicians interpreted chest radiographs of 636 asbestos-exposed workers[8)]. Their reports of an overall kappa value of 0.29 for small opacity profusion was considerably lower than the American study and ours, while 0.42 for pleural lesions was within the range of our findings. Another American study[7)] evaluated 79,185 matched readings by A and B Readers from a coal workers' surveillance program; moderate agreement was seen only on the size of large opacities (kappa value 0.50). (A Readers and B Readers are certified by the NIOSH of the USA. A physician can achieve A Reader status by attending a NIOSH-authorized course on the ILO classification system or submitting radiographs to the NIOSH with ILO classifications for review. To become a B Reader, a physician must pass a rigorous competency-based examination and maintaining B Reader status requires passing the recertification examination every 5 years. In the referenced study[7)], B Readers classified more pneumoconiosis chest radiographs than A Readers did.) The authors concluded that the differences between readers in terms of training in the use of ILO classification and reading experiences were the likely reasons for the observed unsatisfactory agreement in classifying pleural changes (kappa value 0.16) and small opacity profusion (kappa value 0.24)[7)]. In addition to the observers' characteristics, we suggested that the differences in study designs (including the number of radiographs and readers), the defined classifications for studied conditions, and the quality of chest radiographs being classified might have also contributed to the varying degree of inter-observer agreement found across studies.

Specialty training affects the level of diagnostic accuracy and hence the degree of agreement in classifying chest radiographs for pneumoconiosis. A past study reported the existence of differences in diagnostic capability between specialties in reviewing chest radiographs[20)]. Our observation of the radiologists' group showing the highest performance, followed by the pulmonologists' group and the other specialties, also support this (Fig. 1; Table 3). Different physicians may have different thresholds for judging a chest radiograph between normal and abnormal. They may also have differing abilities to observe and recognize radiological appearances of pneumoconiotic lesions. The training to become a radiologist or a pulmonologist differs from that of other specialties. Also, radiologists and pulmonologists may have reviewed many more chest radiographs in routine work than physicians of other specialties. In our study, we observed that radiologists made up the highest proportion of "Country 4" and pulmonologists formed the majority in "Country 2" (Table 1); this uneven representation of specialties between countries was the likely source for differences found between countries.

Physicians' working years, as determined by years after graduation, did not ensure for a better agreement or higher accuracy. We observed better performance from the recent graduates (i.e., five or fewer years working after graduation) (Tables 2 and 3; Fig. 1). Uneven distribution of radiologists and pulmonologists between groups in our study might be one possible explanation for this observation. One previous study noted that to achieve high-level expertise in radiology requires a combination of radiology-specific training and deliberate practice, rather than an absolute number of working years[21)]. Other reasons might be related to the nature of the AIR Pneumo training program. Being younger, recently graduated physicians might be able to absorb more information during the two days of intensive training than their seniors. Also, recent graduates would still be familiar with the time-limited examination environment and manage to produce better results.

Physicians' familiarity with the ILO classification and standard radiographs likely plays a significant role in the reading performance of our physicians. A past study suggested that the number of reviewed chest radiographs also contributed to the poor agreement between A Readers and B Readers[7)]. However, we observed that relatively more numbers in reviewed pneumoconiosis chest radiographs appeared to be of no assistance to better observer agreement or higher accuracy in our physicians. A possible explanation might be that our physicians are not using ILO classification or the standard radiographs in their routine work. And thus, their reading experiences could not provide superior results in a test that required the ILO classification. Although we had not tested for it, our physicians' levels of understanding of the ILO classification might vary, contributing to the variation seen among groups.

Our physicians' diagnostic accuracy for pleural plaques appeared less satisfactory compared with parenchymal lesions. This finding was very similar to that observed in the U.S. B Reader program. Studies reported that physicians generally classify pleural changes poorly compared with parenchymal lesions, and this nature was the same for physicians who passed or failed the B Reader examinations[22, 23)]. Without specific radiological expertise, the detection of pleural plaques in a chest radiograph becomes challenging. Pleural plaques are irregular, circumscribed lesions on the parietal pleura. Radiographically, they appear as discrete areas of pleural thickening and are barely visible in some cases[2)]. In posteroanterior chest radiographs, shadows of anatomical structures (e.g., subcostal fat, serratus anterior muscles) or pleural thickening secondary to medical conditions (e.g., trauma, infection) may mimic plaques, and distinguishing them required a good knowledge of local anatomy and considerable experience[2, 24, 25)]. A systematic review reported high false-negative and varying false-positive rates in diagnosing pleural plaques on a chest radiograph[24)]. In a recent chest radiograph reading trial involving four readers with different clinical and radiography interpretation experiences (one B Reader and three AIR Pneumo-certified physicians), the investigators reported a lower detection rate for pleural plaques compared with those for parenchymal lesions[26)]. They also demonstrated that the detection rate varied among readers, with the most experienced one showing the highest rate. A similar trend was also seen in a study using surveillance data, where B Readers having far greater experiences in diagnosing pneumoconiosis identified more pleural plaques than A Readers did[7)]. In the present study, our physicians, except the radiologists, showed a lower accuracy in identifying pleural plaques when compared with those of parenchymal lesions, indicating specific training is required to develop diagnostic accuracy and improve agreement in the diagnosis of pleural plaques.

Accurate diagnosis and reporting from physicians are vital to the success of screening programs and disease prevention. The ILO/WHO's Global Program for the Elimination of Silicosis (GPES), aiming to eliminate new cases of silicosis from all workplaces by 2030, set its strategy on early detection of diseases through surveillance along with dust exposure control[27)]. Similarly, the WHO's Global campaign for the elimination of asbestos-related disease works through improving early diagnosis and establishing registries of people with past and/or current exposures along with other primary preventive measures[28)]. A recent article reported the worldwide occurrence of increasing incidence of pneumoconiosis for the last three decades. Of the 60,055 incident cases in 2017, more than half occurred in Asia: 32,305 cases in China and 5,160 cases in India[29)]. Moreover, as the importation and use of asbestos in developing Asian countries has been continuing, a substantial number of people may have been exposed to asbestos occupationally and non-occupationally[30)]. In these circumstances, our findings have several important occupational and public health implications. First, we reported the degree of inter-observer agreement and sources for variation in classifying pneumoconiotic lesions among Asian physicians taking AIR Pneumo examination. The awareness of variability allows a careful comparison of results between different studies and knowledge of the source enables us to recommend measures to correct the variations. Second, we observed a low-level diagnostic accuracy and poor agreement in classifying radiographs for pleural plaques. Pleural plaques indicate past exposure to asbestos[31)]; in most cases, they are asymptomatic and often identified as incidental chest radiographic findings[32)]. Attending physician's familiarity with the radiological appearance of pleural plaques is central to their identification. The ILO standards radiographs illustrate a spectrum of radiological appearances seen in all types of pneumoconiotic lesions[4)], the use of which permits physicians' familiarity with radiological appearances of pneumoconiosis, and thereby, improves diagnostic accuracy, especially for less experienced physicians[33)]. Training in the use of the ILO classification, such as that provided by the AIR Pneumo, might promote physicians' reading skill further[34)].

This study has several limitations. First, we used data derived from examinations. Participants might expect more radiographs showing signs of pneumoconiosis and assess them in a manner different from their routine work. However, we believed that the participants' enthusiasm and compliance with the standard assessment procedure made the data featured their actual performance in applying the ILO classification. Second, since our physicians have a common interest in pneumoconiosis, findings in this study may not necessarily represent the performance of Asian physicians in general. However, it should be noted that our physicians are grossly representing the physician population in pneumoconiosis screening in their respective countries. Third, we do not have information on the requirements of specialty training in each country. But we believe these might differ between specialties and between countries. We suggested the uneven specialty representation within each country requires careful interpretation of individual country results. Fourth, the different number of readers among the groups studied might affect the estimated kappa coefficients.

## CONCLUSION

Reviewing chest radiographs using the ILO classification is the current international standard in screening for pneumoconiosis. Asian physicians taking the AIR Pneumo examination were better at classifying parenchymal lesions than pleural plaques using the ILO classification. The degree of inter-observer agreement differed among countries, and this difference was associated with a physician's specialty training background. Specific training on the use of the ILO classification, as provided by the AIR Pneumo, and continuing practice would improve diagnostic accuracy and lessen observer variability.

**Acknowledgments:** The AIR Pneumo project consumed great amount of work, research, and dedication. Still, it's training and examination programs would not have been possible without supports from many individuals and organizations. Therefore, we would like to extend our sincere gratitude to all of them, but the list here is not exhaustive. The authors gratefully acknowledge (1) the Supporting Bodies for AIR Pneumo program, which includes the Scientific Committee on Respiratory Diseases, International Commission on Occupational Health (ICOH), Asian Pacific Society of Respirology, Japan Society for Occupational Health (JSOH), and University of Fukui (Fukui, Japan); (2) the co-organizers, which include the Thailand Association of Occupational and Environmental Diseases and the Central Chest Institute of Thailand, Bureau of Occupational and Environmental Diseases, Ministry of Public Health (Thailand); (3) the sponsors, which include Thailand Workmen's Compensation Fund and Social Security Office (Thailand); (4) personal advisors, which include Dr. Prahalad K. Sishodiya (India), Kazutaka Kogi (Japan), and Yoshiharu Aizawa (Japan), G.R. Wagner (USA), Dr. Tran Anh Thanhh (Vietnam), Dr. Khuong Van Duy (Vietnam); and (5) the volunteer experts who devoted their time and knowledge in the training programs as lecturers.

**Authors' contribution:** All authors contributed toward data collection and reviewed and approved this manuscript. NA J-P: Writing original draft, data curation, data analysis, review & editing. NS: Writing original draft, data curation, data analysis, review & editing. ADS: Data curation, review & editing. ES: Data curation, review & editing. MM: Data curation, review & editing. ST: Data curation, review & editing. SL: Data curation, review & editing. PS: Data curation, review & editing. SS: Data curation, review & editing. ND: Data curation, review & editing. EA: Data curation, review & editing. JEP: Data curation, review & editing.

KGH: Data curation, review & editing. HK: Data curation, review & editing. TT: Data curation, review & editing. YK: Data curation, review & editing.

**Conflict of interest:** None declared.

## REFERENCES

1) Chong S, Lee KS, Chung MJ, Han J, Kwon OJ, Kim TS (2006) Pneumoconiosis: comparison of imaging and pathologic findings. Radiographics **26**, 59–77.

2) Norbet C, Joseph A, Rossi SS, Bhalla S, Gutierrez FR (2015) Asbestos-related lung disease: a pictorial review. Curr Probl Diagn Radiol **44**, 371–82.

3) Wagner GR (1996) Screening and surveillance of workers exposed to mineral dust. World Health Organization, Geneva, Switzerland.

4) ILO Guidelines for the use of the ILO international classification of radiographs of pneumoconioses. Revised edition 2011. International Labour Office, Geneva, Switzerland.

5) J-P NA, Imanaka M, Suganuma N (2017) Japanese workplace health management in pneumoconiosis prevention. J Occup Health **59**, 91–103.

6) J-P NA, Suganuma N (2020) Quality assurance in reading radiographs for pneumoconiosis: AIR Pneumo program. ASEAN-JR **21**, 73–84.

7) Halldin CN, Blackley DJ, Petsonk EL, Laney AS (2017) Pneumoconioses radiographs in a large population of U.S. coal workers: variability in A Reader and B Reader classifications by using the International Labour Office classification. Radiology **284**, 870–6.

8) Ochsmann E, Carl T, Brand P, Raithel HJ, Kraus T (2010) Inter-reader variability in chest radiography and HRCT for the early detection of asbestos-related lung and pleural abnormalities in a cohort of 636 asbestos-exposed subjects. Int Arch Occup Environ Health **83**, 39–46.

9) Tamura T, Kusaka Y, Suganuma N, Suzuki K, Subhannachart P, Siriruttanapruk S, Dumavibhat N, Zhang X, Sishodiya PK, Thanh TA, Hering KG, Parker JE, Algranti

E, O'Connor FS, Shida H, Akira M (2018) Assessment of physicians' proficiency in reading chest radiographs for pneumoconiosis, based on a 60-film examination set with two factors constituting eight indices. Ind Health **56**, 382–93.

10) Zhou H, Kusaka Y, Tamura T, Suganuma N, Subhannachart P, Siriruttanapruk S, Dumavibhat N, Zhang X, Sishodiya PK, K. VD, Hering KG, Parker JE, Algranti E, Fedotov I, Shida H, Akira M (2012) The 60-film set with 8-index for examining physicians' proficiency in reading pneumoconiosis chest X-rays. Ind Health **50**, 84–94.

11) Gamble JF, Hessel PA, Nicolich M (2004) Relationship between silicosis and lung function. Scand J Work Environ Health **30**, 5–20.

12) Klein D (2016) KAPPAETC: Stata module to evaluate interrater agreement. revised 06 Jan 2019. https://ideas.repec.org/c/boc/bocode/s458283.html. Accessed October 25, 2021.

13) Gisev N, Bell JS, Chen TF (2013) Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Social Adm Pharm **9**, 330–8.

14) Linden A (2006) Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J Eval Clin Pract **12**, 132–9.

15) Youden WJ (1950) Index for rating diagnostic tests. Cancer **3**, 32–5.

16) Amandus HE, Pendergrass EP, Dennis JM, Morgan WK (1974) Pneumoconiosis: inter-reader variability in the classification of the type of small opacities in the chest roentgenogram. Am J Roentgenol Radium Ther Nucl Med **122**, 740–3.

17) Rossiter CE (1972) Initial repeatability trials of the UICC-Cincinnati classification of the radiographic appearances of pneumoconioses. Br J Ind Med **29**, 407–19.

18) Takashima Y, Suganuma N, Sakurazawa H, Itoh H, Hirano H, Shida H, Kusaka Y (2007) A flat-panel detector digital radiography and a storage phosphor computed radiography: screening for pneumoconioses. J Occup Health **49**, 39–45.

19) Halldin CN, Petsonk EL, Laney AS (2014) Validation of the International Labour Office digitized standard images for recognition and classification of radiographs of pneumoconiosis. Acad Radiol **21**, 305–11.

20) Cascade PN, Kazerooni EA, Gross BH, Quint LE, Silver TM, Bowerman RA, Pernicano PG, Gebremariam A (2001) Evaluation of competence in the interpretation of chest radiographs. Acad Radiol **8**, 315–21.

21) Kelly BS, Rainford LA, Darcy SP, Kavanagh EC, Toomey RJ (2016) The development of expertise in radiology: in chest radiograph interpretation, “expert” search pattern may predate “expert” levels of diagnostic accuracy for pneumothorax identification. Radiology **280**, 252–60.

22) Halldin CN, Hale JM, Weissman DN, Attfield MD, Parker JE, Petsonk EL, Cohen RA, Markle T, Blackley DJ, Wolfe AL, Tallaksen RJ, Laney AS (2019) The National Institute for Occupational Safety and Health B Reader certification program-An update report (1987 to 2018) and future directions. J Occup Environ Med **61**, 1045–51.

23) Wagner GR, Attfield MD, Kennedy RD, Parker JE (1992) The NIOSH B Reader certification program. An update report. J Occup Med **34**, 879–84.

24) Clarke CC, Mowat FS, Kelsh MA, Roberts MA (2006) Pleural plaques: a review of diagnostic issues and possible nonasbestos factors. Arch Environ Occup Health **61**, 183–92.

25) Alfudhili KM, Lynch DA, Laurent F, Ferretti GR, Dunet V, Beigelman-Aubry C (2016) Focal pleural thickening mimicking pleural plaques on chest computed tomography: tips and tricks. Br J Radiol **89**, 20150792.

26) Nogami S, J-P NA, Nogami M, Matsui T, Ngatu NR, Tamura T, Kusaka Y, Itoh H, Suganuma N (2020) Radiographic diagnosis of pneumoconioses by AIR Pneumo-trained physicians: comparison with low-dose thin-slice computed tomography. J Occup Health **62**, e12141.

27) Fedotov IA, Eijkemans GJM (2007) The ILO/WHO global programme for the elimination of silicosis (GPES). GOHNET Newsletter.

28) WHO. Asbestos: elimination of asbestos-related diseases. https://www.who.int/news-room/fact-sheets/detail/asbestos-elimination-of-asbestos-related-diseases. Accessed October 25, 2021.

29) Shi P, Xing X, Xi S, Jing H, Yuan J, Fu Z, Zhao H (2020) Trends in global, regional and national incidence of pneumoconiosis caused by different aetiologies: an analysis from the Global Burden of Disease Study 2017. Occup Environ Med **77**, 407–14.

30) Leong SL, Zainudin R, Kazan-Allen L, Robinson BW (2015) Asbestos in Asia. Respirology **20**, 548–55.

31) Wolff H, Vehmas T, Oksa P, Rantanen J, Vainio H (2015) Asbestos, asbestosis, and cancer, the Helsinki criteria for diagnosis and attribution 2014: recommendations. Scand J Work Environ Health **41**, 5–15.

32) Cugell DW, Kamp DW (2004) Asbestos and the pleura: a review. Chest **125**, 1103–17.

33) Fletcher CM, Oldham PD (1951) The use of standard films in the radiological diagnosis of coalworkers' pneumoconiosis. Br J Ind Med **8**, 138–49.

34) Ngatu NR, Suzuki S, Kusaka Y, Shida H, Akira M, Suganuma N (2010) Effect of a two-hour training on physicians' skill in interpreting Pneumoconiotic chest radiographs. J Occup Health **52**, 294–301.

## Figure legends

**Fig. 1.** Accuracy in classifying radiographs for the presence or absence of pneumoconiosis. Average AUC values of physician groups formed by (A) country, (B) specialty, (C) years after graduation, and (D) number of reviewed pneumoconiosis chest radiographs

**Table 1.** Information of the physicians

| Total | Physicians (n=93) | Country 1 (n=6) | 2 (n=54) | 3 (n=10) | 4 (n=23) |
|---|---|---|---|---|---|
| | Number of physicians (%) | | | | |
| **Gender** | | | | | |
| Female | 50 (53.8) | 2 (33.3) | 32 (59.3) | 5 (50.0) | 11 (47.8) |
| Male | 34 (36.6) | 4 (66.7) | 17 (31.5) | 5 (50.0) | 8 (34.8) |
| Missing | 9 (9.7) | 0 | 5 (9.3) | 0 | 4 (17.4) |
| **Specialty** | | | | | |
| Pulmonology | 40 (43.0) | 0 | 38 (70.4) | 1 (10.0) | 1 (4.3) |
| Occupational medicine | 25 (26.9) | 4 (66.7) | 10 (18.5) | 6 (60.0) | 5 (21.7) |
| Public health | 4 (4.3) | 1 (16.7) | 0 | 2 (20.0) | 1 (4.3) |
| Radiology | 15 (16.1) | 0 | 2 (3.7) | 1 (10.0) | 12 (52.2) |
| Missing | 9 (9.7) | 1 (16.7) | 4 (7.4) | 0 | 4 (17.4) |
| **Years after graduation** | | | | | |
| Median (range) | 6 (1–34) | 15 (5–30) | 6 (1–34) | 8.5 (4–23) | 3 (1–34) |
| ≤5 | 37 (39.8) | 1 (16.7) | 21 (38.9) | 1 (10.0) | 14 (60.9) |
| 6–10 | 27 (29.0) | 2 (33.3) | 16 (29.6) | 6 (60.0) | 3 (13.0) |
| ≥11 | 15 (16.1) | 3 (50.0) | 7 (13.0) | 3 (30.0) | 2 (8.7) |
| Missing | 14 (15.1) | 0 | 10 (18.5) | 0 | 4 (17.4) |
| **Number of reviewed pneumoconiosis CXR** | | | | | |
| None | 17 (18.3) | 0 | 12 (22.2) | 3 (30.0) | 2 (8.7) |
| <10 | 41 (44.1) | 2 (33.3) | 27 (50.0) | 3 (30.0) | 9 (39.1) |
| <50 | 20 (21.5) | 2 (33.3) | 8 (14.8) | 4 (40.0) | 6 (26.1) |
| ≥50 | 6 (6.4) | 2 (33.3) | 2 (3.7) | 0 | 2 (8.7) |
| Missing | 9 (9.7) | 0 | 5 (9.3) | 0 | 4 (17.4) |

**Table 2.** Inter-observer agreement in classifying radiographs for pneumoconiosis[a]

| | Small opacity profusion[b] | Small opacity shape[c] | Large opacity size[b] | Presence of pleural plaque[c] |
|---|---|---|---|---|
| | Fleiss' kappa coefficient (95% CI) | | | |
| **Physician overall** | 0.59 (0.52–0.67) | 0.29 (0.23–0.36) | 0.65 (0.55–0.74) | 0.30 (0.20–0.40) |
| **Country** | | | | |
| 1 | 0.50 (0.39–0.61) | 0.18 (0.05–0.32) | 0.57 (0.42–0.72) | 0.34 (0.19–0.49) |
| 2 | 0.59 (0.51–0.67) | 0.26 (0.20–0.32) | 0.66 (0.57–0.75) | 0.25 (0.16–0.34) |
| 3 | 0.47 (0.38–0.55) | 0.21 (0.13–0.30) | 0.55 (0.40–0.70) | 0.31 (0.20–0.42) |
| 4 | 0.73 (0.66–0.80) | 0.56 (0.48–0.65) | 0.69 (0.59–0.79) | 0.55 (0.42–0.68) |
| **Specialty** | | | | |
| Pulmonology | 0.62 (0.54–0.69) | 0.26 (0.20–0.31) | 0.69 (0.61–0.77) | 0.29 (0.19–0.38) |
| Occupational medicine | 0.53 (0.45–0.61) | 0.28 (0.20–0.37) | 0.56 (0.44–0.68) | 0.26 (0.16–0.35) |
| Public health | 0.51 (0.39–0.64) | 0.12 (0.02–0.22) | 0.56 (0.38–0.75) | 0.30 (0.12–0.48) |
| Radiology | 0.69 (0.61–0.77) | 0.54 (0.45–0.64) | 0.74 (0.64–0.83) | 0.58 (0.44–0.71) |
| **Years after graduation** | | | | |
| ≤5 | 0.67 (0.60–0.75) | 0.39 (0.32–0.46) | 0.72 (0.63–0.80) | 0.39 (0.27–0.51) |
| 6–10 | 0.52 (0.44–0.61) | 0.21 (0.16–0.27) | 0.59 (0.48–0.70) | 0.26 (0.17–0.35) |
| ≥11 | 0.53 (0.45–0.61) | 0.28 (0.20–0.36) | 0.55 (0.43–0.67) | 0.24 (0.14–0.34) |
| **Number of reviewed pneumoconiosis CXR** | | | | |
| None | 0.55 (0.48–0.62) | 0.23 (0.15–0.31) | 0.61 (0.52–0.70) | 0.22 (0.14–0.30) |
| <10 | 0.63 (0.55–0.71) | 0.32 (0.26–0.39) | 0.68 (0.58–0.78) | 0.33 (0.22–0.43) |
| <50 | 0.56 (0.47–0.64) | 0.31 (0.24–0.38) | 0.60 (0.46–0.73) | 0.29 (0.17–0.41) |
| ≥50 | 0.53 (0.42–0.64) | 0.23 (0.11–0.36) | 0.68 (0.57–0.79) | 0.34 (0.19–0.49) |

a= Computation included the readings of 40 radiographs (9 boundary cases and 31 radiographs with small opacities) for "small opacity shape"; included readings of all 60 radiographs for the others. b= Weighted kappa coefficient. c= Unweighted kappa coefficient.

All kappa coefficients were significant at $p<0.001$.

Interpretation of kappa coefficients: <0.2 = poor, 0.21–0.4 = fair, 0.41–0.6 = moderate, 0.61–0.8 = good, and 0.81–1.0 = almost perfect agreement.

**Table 3.** Comparison between physician groups for accuracy in classifying radiographs for pneumoconiosis

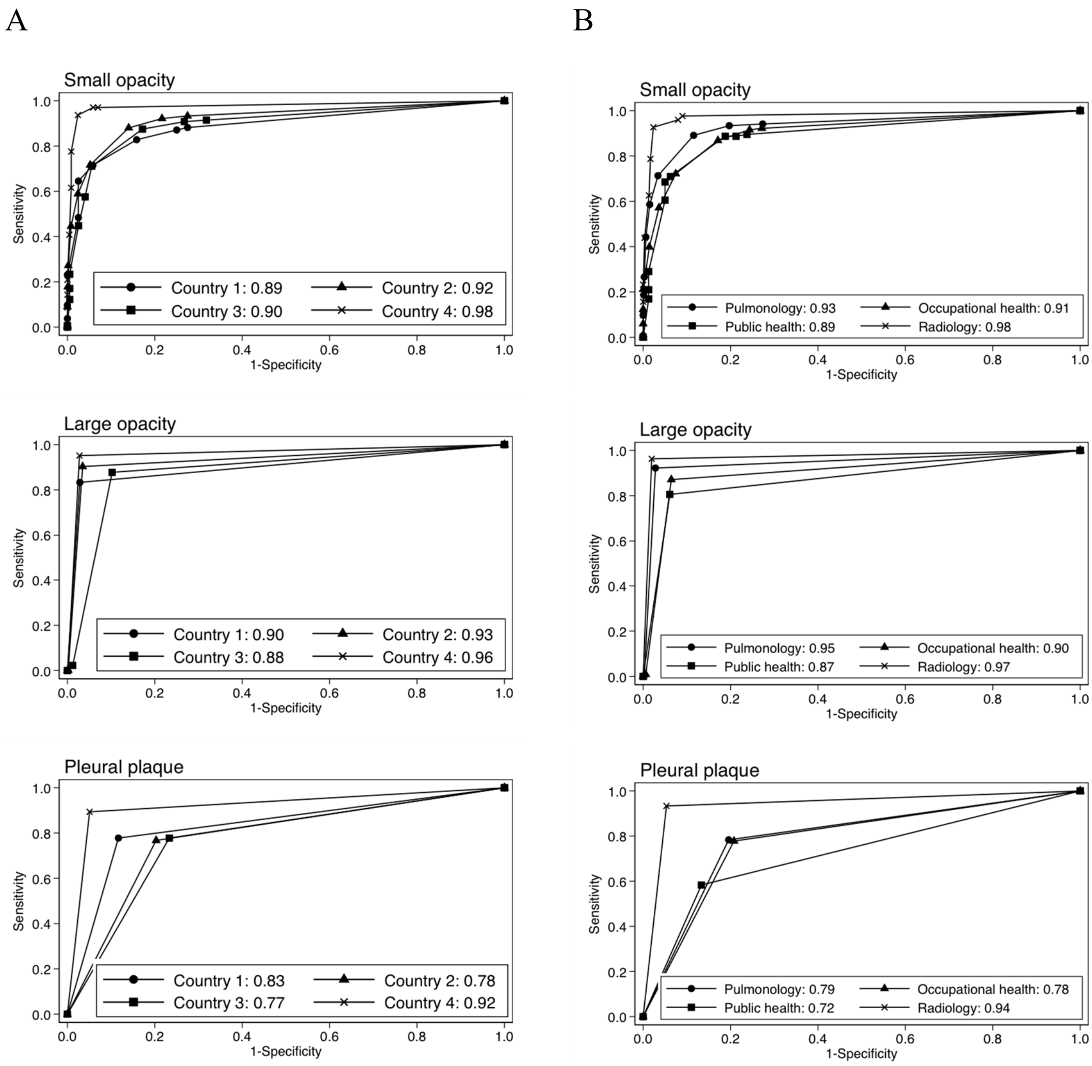
| | Physicians | Small opacity | Small opacity shape | Large opacity | Pleural plaque |
|---|---|---|---|---|---|
| | **Number** | **Accuracy score[a], Median (Interquartile range)** | | | |
| **Physician overall** | 93 | 76.8 (29.3) | 83.3 (25) | 88.9 (23.3) | 61.1 (25.5) |
| **Country** | | | | | |
| 1 | 6 | 63.7 (13.5)*** | 79.2 (12.5) | 74.1 (27.1) | 68.3 (15.6) |
| 2 | 54 | 73.9 (30.8)*** | 75 (20.8)*** | 88.9 (20.9) | 56.1 (25.5)*** |
| 3 | 10 | 62.5 (16.8)*** | 66.7 (20.8)*** | 77.2 (19.5)** | 55.0 (15.6)*** |
| 4 (Reference) | 23 | 91.8 (12.9) | 95.8 (12.5) | 97.6 (13.5) | 85.6 (18.9) |
| Kruskal-Wallis test | | *p*<0.001 | *p*<0.001 | *p*=0.012 | *p*<0.001 |
| **Specialty** | | | | | |
| Radiology (Reference) | 15 | 91.8 (25.0) | 100 (16.7) | 100 (13.5) | 85.6 (18.9) |
| Pulmonology | 40 | 77.1 (26.0)* | 79.2 (22.9)** | 95.1 (16.0) | 58.3 (25.0)*** |
| Occupational health | 25 | 67.1 (35.8)** | 75 (16.7)** | 81.6 (30.6)** | 60.0 (15.5)*** |
| Public health | 4 | 74.6 (32.2) | 70.8 (31.2)* | 72.4 (12.9)* | 47.8 (13.3)*** |
| Kruskal-Wallis test | | *p*=0.005 | *p*=0.002 | *p*=0.003 | *p*<0.001 |
| **Years after graduation** | | | | | |
| ≤5 (Reference) | 37 | 87.1 (19.7) | 87.5 (20.8) | 97.6 (11.1) | 75.6 (27.8) |
| 6–10 | 27 | 70.0 (34.4)** | 66.7 (29.2)*** | 84.0 (20.9)** | 55.6 (38.9)** |
| ≥11 | 15 | 67.1 (27.6)** | 83.3 (12.5) | 85.4 (16.0)* | 57.8 (15.6)** |
| Kruskal-Wallis test | | *p*<0.001 | *p*= 0.002 | *p*=0.005 | *p*=0.001 |
| **Number of reviewed pneumoconiosis CXR** | | | | | |
| None (Reference) | 17 | 70 (27.1) | 75 (25) | 88.9 (18.4) | 53.3 (25.6) |
| <10 | 41 | 83.9 (26.8) | 83.3 (25) | 95.1 (16.0) | 67.8 (33.3) |
| <50 | 20 | 75.2 (30.9) | 83.3 (22.9) | 82.8 (25.3) | 61.1 (28.9) |
| ≥50 | 6 | 71.2 (26.4) | 70.8 (29.2) | 96.3 (8.7) | 67.8 (26.7) |
| Kruskal-Wallis test | | *p*=0.10 | *p*=0.206 | *p*=0.113 | *p*=0.139 |

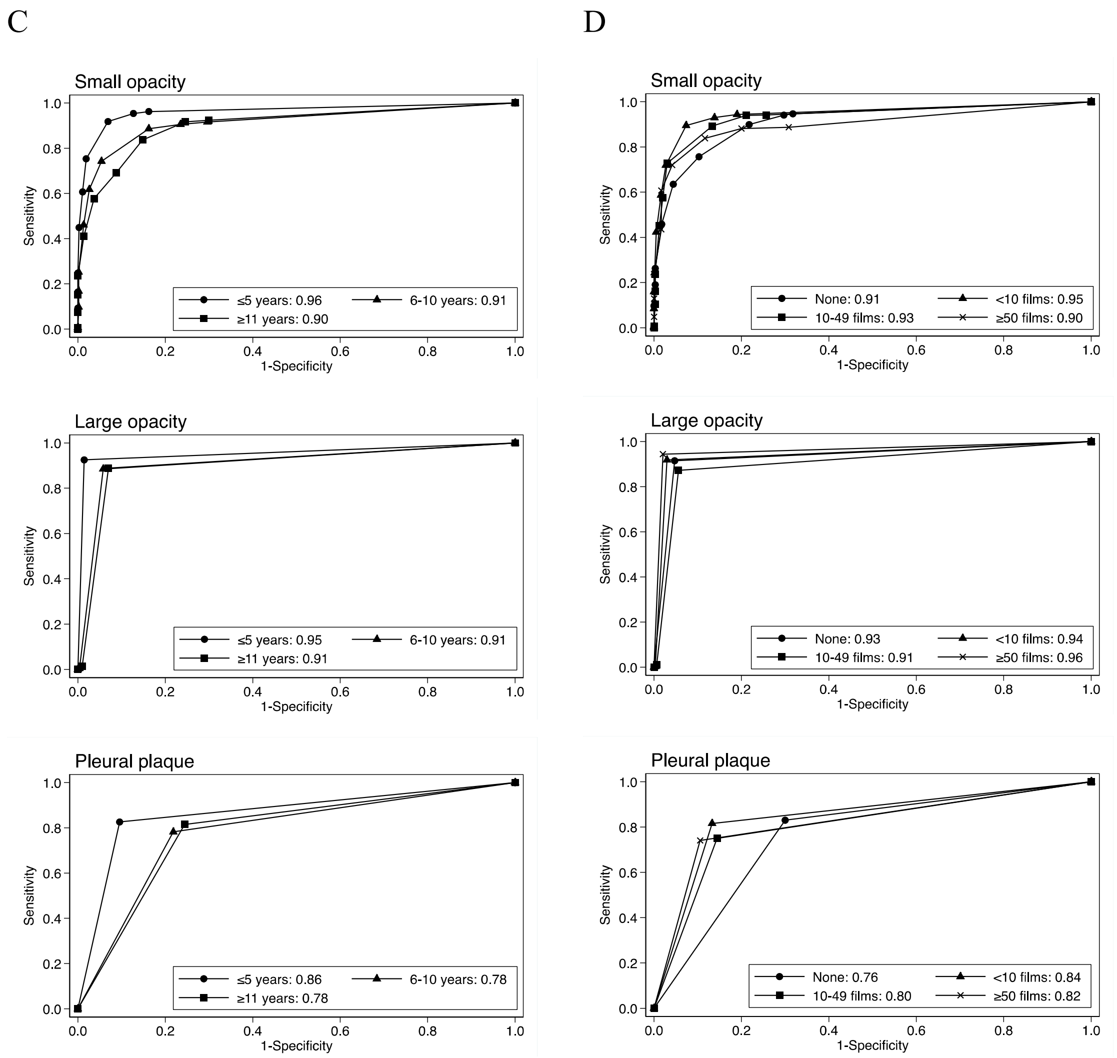
a= Accuracy scores are calculated as Youden's J index x 100, except for "small opacity shape". Scores for "small opacity shape" are percent agreement with experts' classification of small opacities as rounded or irregular.

Reference = reference group in Bonferroni correction for multiple comparisons

*p* values of Bonferroni correction for multiple comparisons: * *p*<0.05, ** *p*<0.01, *** *p*<0.001.

**Fig. 1**


A
B
Small opacity
Sensitivity
1-Specificity
Country 1: 0.89
Country 2: 0.92
Country 3: 0.90
Country 4: 0.98
Large opacity
Country 1: 0.90
Country 2: 0.93
Country 3: 0.88
Country 4: 0.96
Pleural plaque
Country 1: 0.83
Country 2: 0.78
Country 3: 0.77
Country 4: 0.92
Small opacity
Pulmonology: 0.93
Occupational health: 0.91
Public health: 0.89
Radiology: 0.98
Large opacity
Pulmonology: 0.95
Occupational health: 0.90
Public health: 0.87
Radiology: 0.97
Pleural plaque
Pulmonology: 0.79
Occupational health: 0.78
Public health: 0.72
Radiology: 0.94


C
D
Small opacity
Sensitivity
1-Specificity
≤5 years: 0.96
6-10 years: 0.91
≥11 years: 0.90
Large opacity
≤5 years: 0.95
6-10 years: 0.91
≥11 years: 0.91
Pleural plaque
≤5 years: 0.86
6-10 years: 0.78
≥11 years: 0.78
Small opacity
None: 0.91
<10 films: 0.95
10-49 films: 0.93
≥50 films: 0.90
Large opacity
None: 0.93
<10 films: 0.94
10-49 films: 0.91
≥50 films: 0.96
Pleural plaque
None: 0.76
<10 films: 0.84
10-49 films: 0.80
≥50 films: 0.82